Adaptive Designs
IMS Lecture Notes - Monograph Series (1995) Volume 25

SEQUENTIAL ALLOCATION INVOLVING SEVERAL TREATMENTS

BY D.S. COAD

The University of Sussex

Abstract

A clinical trial model is considered in which $k \geq 2$ treatments are compared and treatment allocation is data-dependent. A sequential procedure for determining the best treatment is investigated that is a natural generalization of the test for two treatments studied by Robbins and Siegmund (1974). It is shown by extensive simulation that the error probability for the procedure is insensitive to the data-dependent allocation rule used. The estimation formulae of Coad (1994) are shown to give good approximations to the bias and variance of estimators of treatment differences.

1. Introduction. Suppose a clinical trial is conducted in which patients can be allocated to one of $k \geq 2$ treatments. The response variable for treatment i at time j, X_{ij} $(j = 0, 1, \ldots)$, is normally distributed with mean μ_i and variance unity. The sequential procedure we shall consider is symmetric with respect to the ordering of the treatments, so properties of the procedure will be invariant under permutations of the means. Thus, although the means are unknown, we shall assume for convenience that $\mu_1 > \mu_2 \geq \mu_3 \geq \cdots \geq \mu_k$. During the trial, a treatment can be eliminated if it does not look promising. At the end

Received October 1992; revised April 1993.
AMS 1991 subject classification. Primary 62F07, 62L05, 62L10, 62L12.
Key words and phrases. Brownian motion, clinical trials, data-dependent treatment allocation, error probabilities, estimation, Gittins index.

of the trial, we wish to choose the treatment that has the highest mean response. Additionally, we wish to estimate the treatment differences $\mu_i - \mu_j$ at the last stage in which both treatments are still in use.

For the case $k = 2$, the above testing problem was studied in detail by Robbins and Siegmund (1974). A sequential probability ratio test was derived for which the test statistics at stage (m, n), that is when m patients have been allocated to Treatment 1 and n to Treatment 2, is

(1)
$$z_{m,n} = \frac{mn}{(m+n)} \left(\bar{x}_{1m} - \bar{x}_{2n} \right),$$

where $\bar{x}_{1m}$ and $\bar{x}_{2n}$ are the sample means on Treatments 1 and 2, respectively. Robbins and Siegmund showed that for a large class of allocation rules, the random sequence $\{z_{m,n}\}$ has the same joint distribution as Brownian motion with drift $\mu_1 - \mu_2$ observed at times $\{mn/(m+n)\}$. This result led to the important finding that the error probability for the test defined by (1) is approximately independent of the allocation rule used. Such a test enables us to seek a data-dependent allocation rule that reduces the number of patients on the inferior treatment relative to pairwise allocation where there are equal number of patients on each treatment.

There has now been substantial progress in the testing problem for $k = 2$: for example, see Coad (1991). But, until recently, the problem of estimation following the test had not been addressed. However, Woodroofe (1989) derived an asymptotic expansion for the distribution function of (1) at the end of the test, in terms of the standard normal distribution function. More recently, Coad (1994) obtained approximations for the bias and variance of the maximum likelihood estimator of $\mu_1 - \mu_2$ upon termination of the test.

Much less attention has been given to the comparison of more than two treatments when data-dependent allocation used. The simplest case is the δ-slippage configuration $\mu_1 - \delta = \mu_2 = \mu_3 = \ldots = \mu_k$, where $\delta > 0$. This case was studied by Turnbull, Kaspi and Smith (1978), who investigated the cases $k = 3$ and $k = 10$. A more detailed study of the case $k > 2$ was carried out by Jennison, Johnstone and Turnbull (1980), the main results of which appeared in Jennison, Johnstone and Turnbull (1982). In its simplest form, the sequential procedure studied by Jennison, Johnstone and Turnbull (1980) is a natural generalization of

the sequential test for $k = 2$. Suppose we are at stage $(n_1,\ n_2,\ \ldots,\ n_k)$ in the trial, that is when n_i patients have been allocated to treatment i for $i = 1,\ 2,\ \ldots,\ k$, and define the test statistics

(2)
$$z_{ij}\left(n_i, n_j\right) = \frac{n_i n_j}{\left(n_i + n_j\right)}\left(\bar{x}_i\left(n_i\right) - \bar{x}_j\left(n_j\right)\right),$$

where $\bar{x}_i\left(n_i\right)$ denotes the sample mean on treatment i. At each stage in the trial, we compute the test statistic (2) for all pairs of treatments that have not been eliminated before that stage. We eliminate any treatment j for which $z_{ij}\left(n_i, n_j\right) \geq b$ for some $i \neq j$, where b is a positive constant.

As indicated earlier, Robbins and Siegmund showed that for $k = 2$ the random sequence $\{z_{m,n}\}$ in (1) behaves like Brownian motion. They also showed that the sequence

$$\left\{z_{m,n} - mn\left(\mu_1 - \mu_2\right)/\left(m+n\right),\ m,\ n = 1,\ 2,\ \ldots\right\}$$

is a martingale. Jennison, Johnstone, and Turnbull (1980) showed that, in general, the Brownian motion and the martingale properties of (2) are not preserved when $k > 2$, but it was pointed out that, for this case, martingales can be constructed that are essentially linear combinations of the test statistics defined by (2). Furthermore, simulation for $k = 10$ suggested that Brownian motion provides a reasonable approximation to the distribution of the random sequence $\left\{z_{ij}\left(n_i, n_j\right)\right\}$.

In this paper, we wish to investigate the case $k > 2$ more thoroughly. We shall first compare the error probabilities for the procedure when using different data-dependent allocation rules. For a given set of parameter values, we would expect these to be similar if Brownian motion is a good approximation to the distribution of the random sequence $\left\{z_{ij}\left(n_i, n_j\right)\right\}$. We shall also compare the numbers of patients on the inferior treatments with those using equal randomization. Finally, the bias and variance formulae of Coad (1994) will be used to approximate the bias and variance of estimators of the treatment differences $\mu_i - \mu_j$.

We begin in Section 2 by defining the criteria we shall use to compare the different methods of allocation. Several data-dependent allocation rules are described in Section 3. In Section 4; simulation results are

presented for $k = 3$ and $k = 5$. Some concluding remarks are made in Section 5.

2. Criteria for comparing allocation rules. Several criteria will be used to assess the performance of different data-dependent allocation rules. We seek rules that are equivalent in terms of choosing the best treatment at the end of the trial. Thus, defining the error probability (EP) to be the probability of eliminating the best treatment, we seek rules that achieve a similar pattern of error probabilities over the k-dimensional parameter space of the means.

It was demonstrated by Bather and Coad (1992) that the use of equal randomization, in which equal numbers of patients are randomized to the surviving treatments in each stage, can substantially reduce the total expected number of patients in the trial (ASN) involving $k > 2$ treatments, as compared with a fixed-sample procedure. The numbers of patients on the inferior treatments are also considerably reduced, and they can be further reduced by using data-dependent allocation. Moreover, in contrast to the case $k = 2$, a reduction in the numbers of patients on the inferior treatments does not necessarily lead to an increase in the total number of patients in the trial. This aspect of data-dependent allocation seems to have received very little attention in the literature. It is also important to note that, for the sequential procedure defined by the test statistics (2), the maximum reduction, relative to equal randomization, in the expected total number of patients on the inferior treatments (ITN) is 50%: see Jennison, Johnstone, and Turnbull (1980).

For Bernoulli responses, a useful measure of the cost to patients in a trial is the expected successes lost (ESL), which is the difference between the expected number of successes when all the patients receive the best treatment and the expected number of successes for the trial. We can define an analogous quantity for normal responses. Let N_i denote the number of patients receiving treatment i for $i = 1, 2, \ldots, k$. Then we define

(3)
$$ESL = \sum_{i=2}^{k} (\mu_1 - \mu_i)\, E\,(N_i)\,.$$

Since the martingale properties of (2) in general are not preserved when $k > 2$, it is difficult to establish an upper bound for the ESL in (3).

However, for equal randomization, by neglecting overshoot of the stopping boundaries, it is easy to see that a useful bound is given by

(4)
$$ESL \leq 2(k-1)b.$$

As we shall see in Section 4, the use of data-dependent allocation, in some cases, can significantly reduce the ESL, without increasing the ASN.

We shall also be interested in estimating the treatment differences $\mu_1 - \mu_j$ by the estimators

(5)
$$\hat{\mu}_{ij} = \bar{x}_i(N) - \bar{x}_j(N),$$

where $N = \min(N_i, N_j)$ is the last stage in the trial at which both treatments i and j are used. Although we lose some information by using only the responses up to stage N, we hope to avoid serious biases which might arise from time trends in the data. Furthermore, the bias and variance of the estimators (5) can be approximated, using the expressions derived by Coad (1994) which neglect all but the two treatments directly concerned.

3. Allocation rules. Several data-dependent allocation rules are studied in this paper. These will now be described. Note that for each rule, one patient is initially allocated to each treatment.

EQUAL RANDOMIZATION RULE (EQUAL). *Randomize in the ratios* 1:1: $\cdots$:1 *to the s surviving treatments at each stage.*

For $k = 2$, the equal randomization rule minimizes the ASN. The second rule we consider was suggested by Jennison, Johnstone, and Turnbull (1980); for the δ-slippage configuration of means, this rule asymptotically minimizes the ASN. As we shall see in Section 4, this rule also significantly reduces the number of patients on the inferior treatments for a wide choice of configurations of the means.

JENNISON, JOHNSTONE & TURNBULL RULE (JJT). *Randomize in the ratios*

$$\sqrt{s-1} : 1 : 1 : \cdots : 1,$$

where s is the number of surviving treatments. Here the largest weight refers to the current best treatment.

An interesting generalization of the JJT rule can be obtained by considering the minimization of a weighted average of the ESL and the ASN for the δ-slippage configuration of means. The derived rule is a generalization of the one proposed by Hayre (1979) for $k = 2$. Suppose we write the weighted average as

(6)
$$a\ ESL + c\ ASN,$$

where

$$ESL = \delta \sum_{i=2}^{k} E(N_i),\ ASN = \sum_{i=1}^{k} E(N_i)$$

and a and c are positive constants. We can regard c as the cost of allocating a patient to any treatment and a as the extra cost of allocation to one of the inferior treatments. By following the argument of Jennison, Johnstone, and Turnbull (1980), it can be shown that the following rule asymptotically minimizes (6).

THE GENERALIZED HAYRE RULE (HAYRE). *Randomize in the ratios*

$$\sqrt{\left(1 + \frac{a}{c}\hat{\delta}\right)(s-1)} : 1 : 1 : \cdots : 1,$$

where $\hat{\delta}$ is the estimated difference between the current best and second best treatment means.

One attraction of the Hayre rule is that we can choose a and c to reflect the importance we attach to reducing the ESL, as opposed to the ASN: increasing a relative to c leads to a more adaptive rule. Of course, putting a equal to zero leads to the JJT rule. For practical use, since δ will be unknown, we replace it by the estimated mean difference between the best and second best treatments.

The next rule considered is a generalization of one proposed by Coad (1991). Although this is a complicated rule, it is highly adaptive, and the results in the next section will demonstrate how much we can reduce the ESL, while at the same time highlighting the substantial increase in the ASN over that for the equal randomization rule.

THE GITTINS RULE (GITTINS). *For a discount factor $\alpha \in (0,1)$ and given independent normal priors for the μ_i, Gittins index for treatment i is defined as*

$$\sup_{\tau>0} \frac{E\left\{\sum_{j=0}^{\tau-1} \alpha^j X_{ij}\right\}}{E\left\{\sum_{j=0}^{\tau-1} \alpha^j\right\}},$$

where τ is any stopping time depending on the responses from treatment i.

The allocation rule is based on these indices which can be computed at each stage for normal response variables. Let r be a fixed constant, with $r \geq 1$. If

$$\min_i \{n_i^r\} < \max_j \{n_j\},$$

where i and j range over the surviving treatments, the next patient is randomly allocated to one of the treatments attaining $\min\{n_i^r\}$; otherwise, the next patient is allocated to the treatment that currently has the largest Gittins index, randomizing in the case of ties.

Observe that $r = 1$ indicates equal allocation, while r close to two, say, produces a highly adaptive rule. The choice of α depends on the

ASN, the total expected number of patients in the trial. Now, the discount sequence $(1, \alpha, \alpha^2, \ldots)$ corresponds to a geometric stopping time with mean $(1-\alpha)^{-1}$. Thus, $\alpha = 0.99$ corresponds approximately to an ASN of 100. Note that, although several authors have investigated the use of Gittins indices for clinical trials involving two treatments, results of their use in comparing more than two treatments have not previously been reported.

The last rule we consider is a simple modification of the equal randomization rule with increasing weights given to the better treatments. One of the advantages of this rule is its simplicity but, as we shall see in Section 4, its performance is similar to the generalized Hayre rule.

UNEQUAL RANDOMIZATION RULE (UNEQUAL). *Randomize in the ratios* $2^{s-1} : 2^{s-2} : \cdots : 2 : 1$, *where* s *is the number of surviving treatments.*

4. Simulation results.

4.1. General. For the sequential procedure described in Section 1, 10,000 simulations were carried out for each allocation rule and set of parameter values. Values for the Gittins index are given in Table 1 of Gittins (1989) for a selection of sample sizes; for other sample sizes, approximate values for the indices were obtained by interpolation. To choose a value for b, we used the result for $k = 2$ that a no-overshoot approximation to the error probability is $\{1 + \exp(2\delta b)\}^{-1}$, where $\delta = \mu_1 - \mu_2$. In this paper, we have taken $b = 6$, which for $k = 2$ and $\delta = 0.25$ gives an approximate error probability of 0.05. For later reference, the equivalent fixed-sample procedure requires about 75 patients on each treatment. For the generalized Hayre rule, we have taken $a = 1.0$ and $c = 0.1$. These values indicate that allocation to an inferior treatment is ten times more costly than allocation to the best treatment. For the Gittins rule, the discount factor $\alpha = 0.99$ and $r = 1.5$.

4.2. Evaluation of properties. We wish to ensure that the error probabilities for the different allocation rules are roughly the same. Table 1 gives the error probabilities, the ESL and the expected sample sizes for the case $k = 3$.

It is clear that the error probabilities are insensitive to the allocation rule used. Note that the standard errors of the estimated error probabilities are approximately 0.0007 for an estimate of 0.005 and approximately 0.003 for an estimate of 0.1. By comparing the five rules,

TABLE 1

Simulation results when $k = 3$: the order of the figures is Equal, JJT, Gittins, Hayre and Unequal.

μ_1	1.0	1.0	1.0	1.0	1.0	1.0	1.0
μ_2	0.0	0.5	0.5	0.75	0.75	0.875	0.875
μ_3	0.0	0.0	0.5	0.5	0.75	0.75	0.875
	0.0000	0.0012	0.0036	0.0407	0.0736	0.1878	0.2713
	0.0000	0.0014	0.0038	0.0438	0.0750	0.1940	0.2688
EP	0.0000	0.0024	0.0028	0.0457	0.0693	0.1789	0.2692
	0.0000	0.0018	0.0031	0.0426	0.0683	0.1819	0.2694
	0.0000	0.0021	0.0043	0.0433	0.0723	0.1863	0.2701
	26.41	25.73	25.53	23.63	22.77	18.91	16.22
	23.94	23.94	23.69	22.45	22.07	18.54	15.95
ESL	17.40	17.38	17.16	18.49	20.25	18.36	18.18
	18.57	19.91	20.84	20.60	21.25	17.83	
	19.72	19.96	20.60	20.00	20.94	17.77	16.20
							15.90
	15.92	26.20	31.08	48.19	56.69	71.73	78.37
	17.01	27.39	33.13	49.48	58.87	71.81	78.59
$E(N_1)$	30.71	58.94	76.18	109.73	129.76	140.94	148.27
	31.17	40.87	46.76	60.78	70.38	79.94	86.95
	23.90	35.99	44.10	61.52	72.42	84.37	89.78
	13.21	25.80	25.39	46.62	45.69	67.13	64.54
	11.93	24.72	23.83	45.25	44.12	65.72	63.86
$E(N_2)$	8.70	17.32	17.30	40.47	40.75	75.84	71.74
	9.35	20.67	20.75	42.13	42.58	64.30	65.36
	9.85	21.17	20.60	41.87	41.72	65.75	63.05
	13.21	12.83	25.67	23.96	45.38	42.08	65.21
	12.01	11.58	23.55	22.28	44.17	41.31	63.70
$E(N_3)$	8.70	8.72	17.02	16.74	40.25	35.52	73.74
	9.22	9.57	20.92	20.14	42.43	39.16	64.27
	9.87	9.37	20.60	19.07	42.02	38.19	64.12
	41.84	64.84	82.14	118.76	147.77	180.94	208.12
	40.95	63.69	80.51	117.02	147.17	178.83	206.16
ASN	48.11	84.97	110.50	166.93	210.76	252.31	293.74
	49.74	71.12	88.43	123.05	155.39	183.40	216.57
	43.62	66.54	85.30	122.47	156.16	183.31	216.95

we see that the Gittins rule is most successful in reducing the ESL. However, its ASN is substantially larger than for equal randomization. The standard errors of the estimated ESL are no more than 1%. considering the JJT rule, it is clear that both the ESL and ASN are reduced compared with equal randomization. This reflects the asymptotic optimality property of the JJT rule noted in Section 3. The generalized Hayre and unequal randomization rules can be regarded as compromises between the JJT rule and the Gittins rule. Indeed, these rules have similar performance characteristics: there is a significant reduction in the number of patients on the inferior treatments compared with equal randomization, and moreover, there is only a small increase in the ASN.

For the equivalent fixed-sample procedure, the ASN is 225. So, where there are large differences between the treatment means, there

is a savings in the ASN of about 80%. The maximum ESL for the fixed-sample procedure is 150, which compares with an approximate maximum of 26 in the table for equal randomization: see inequality (4).

The corresponding results for the case $k = 5$ are given in Table 2. These highlight more clearly the same conclusions. Again, for comparison, the ASN and ESL for the equivalent fixed-sample procedure are 375 and 300, respectively. The maximum ESL for equal randomization is about 52. Even for small differences between the treatment means, there is a saving in the ASN of about 15% when using equal randomization, as opposed to the fixed-sample procedure.

4.3. Results for estimation. We now compare the bias and variance of estimators of the form (5). As indicated in Section 2, for $k = 2$, approximations for the bias and variance can be derived by considering the sequential test in continuous time: see Coad (1994) for details. These approximations can also be used for $k > 2$, by neglecting all but the two treatments directly concerned. The values included in this paper are for the approximations after a correction for overshoot has been made. Note that for the equivalent fixed-sample procedure, the bias is zero and the variance is 0.0267.

Table 3 gives the bias and variance of estimators of treatment differences when $k = 3$. For estimators of differences involving Treatment 1, the biases are insensitive to the allocation rule used, and in general, the approximations work well. Note that the standard errors are between 0.003 and 0.005. When there are small differences between the treatment means, the true biases for the more adaptive rules tend to be over estimated by the approximations. This may be due to the trial being longer for these rules, as shown in Table 1. The agreement is poor between the simulated and approximate biases of the estimator of $\mu_2 - \mu_3$.

From Table 3b, we see that the variances are also insensitive to the allocation rule used, and again the approximations work well for differences involving Treatment 1. Here the standard errors are between 1% and 2%. The results for the variance of the estimators of $\mu_2 - \mu_3$ indicate that the approximation is fairly accurate except when the mean for Treatment 1 is much higher than for the inferior treatments.

Similar results were obtained for the case $k = 5$. For the most part, these confirm the conclusions for the case $k = 3$. However, one feature evident from the former is that the approximations for the bias of es-

TABLE 2

Simulation results when $k = 5$: the order of the figures is Equal, JJT, Gittins, Hayre and Unequal.

μ_1	1.0	1.0	1.0	1.0	1.0	1.0	1.0
μ_2	0.0	0.5	0.5	0.75	0.75	0.875	0.875
μ_3	0.0	0.5	0.5	0.75	0.75	0.875	0.875
μ_4	0.0	0.0	0.5	0.5	0.75	0.75	0.875
μ_5	0.0	0.0	0.5	0.5	0.75	0.75	0.875
	0.0001	0.0038	0.0073	0.0768	0.1243	0.3013	0.4068
	0.0000	0.0041	0.0065	0.0766	0.1210	0.2935	0.4000
EP	0.0000	0.0036	0.0057	0.0728	0.0995	0.2667	0.3569
	0.0000	0.0036	0.0065	0.0749	0.1141	0.2795	0.3764
	0.0000	0.0039	0.0061	0.0717	0.1142	0.2814	0.3811
	52.40	50.63	50.16	45.79	44.01	35.64	29.60
	43.59	44.38	44.75	42.19	41.80	33.99	29.57
ESL	34.74	35.00	34.71	36.90	40.43	35.60	35.39
	36.18	39.37	40.61	40.66	41.38	34.51	30.30
	37.61	38.22	40.53	39.39	41.35	34.77	31.67
	17.52	31.13	36.22	57.61	66.80	80.27	81.69
	21.61	35.02	41.96	62.07	70.73	82.11	84.72
$E(N_1)$	37.18	77.53	95.46	136.88	160.22	164.22	166.24
	39.73	50.82	58.07	75.10	82.85	92.72	94.79
	34.24	49.93	61.37	81.85	94.73	103.37	106.75
	13.13	25.44	24.81	45.78	44.18	64.46	59.55
	10.89	22.93	22.33	43.62	41.74	61.75	59.52
$E(N_2)$	8.68	17.69	17.51	41.08	40.73	74.06	70.97
	9.03	20.42	20.20	42.88	41.66	64.15	60.27
	9.34	21.02	20.21	43.79	40.92	67.72	64.05
	13.06	25.61	25.16	43.67	44.15	65.06	59.49
	10.90	23.12	22.36	44.00	41.90	62.48	58.75
$E(N_3)$	8.70	17.61	17.45	41.04	40.79	75.95	70.62
	9.03	20.76	20.09	42.74	41.44	65.19	61.00
	9.43	21.00	20.11	43.33	41.39	67.86	64.04
	13.09	12.60	25.14	22.94	44.08	38.96	58.72
	10.89	10.69	22.44	20.18	41.97	36.76	59.52
$E(N_4)$	8.66	8.66	17.13	16.30	40.33	33.84	70.80
	9.10	9.42	20.44	19.16	41.11	36.57	60.68
	9.38	8.66	20.43	17.57	41.34	35.69	61.79
	13.12	12.50	25.20	22.91	43.63	38.84	59.01
	10.92	10.66	22.39	20.39	41.57	37.07	58.80
$E(N_5)$	8.71	8.69	17.33	16.44	39.88	33.54	70.70
	9.02	9.37	20.50	19.35	41.33	36.80	60.47
	9.46	8.55	20.31	17.65	41.73	35.61	63.44
	69.92	107.28	136.54	194.92	242.84	287.59	318.45
	65.20	102.42	131.47	190.26	237.92	280.17	321.31
ASN	71.92	130.18	164.88	251.72	321.94	381.61	449.34
	75.91	110.78	139.29	199.24	248.39	295.43	337.21
	71.85	109.17	142.42	204.19	260.11	310.24	360.08

TABLE 3A

Results for bias of $x_i(N) - x_j(N)$, *where* $N = \min(N_i, N_j)$, *when* $k = 3$: *upper figures in order are simulated values for Equal, JJT, Gittins, Hayre and Unequal; lowest figure is approximate value.*

μ_1	1.0	1.0	1.0	1.0	1.0	1.0	1.0
μ_2	0.0	0.5	0.5	0.75	0.75	0.875	0.875
μ_3	0.0	0.0	0.5	0.5	0.75	0.75	0.875
	0.1524	0.1535	0.1576	0.1400	0.1307	0.0861	0.0851
	0.1488	0.1531	0.1489	0.1323	0.1294	0.0875	0.0812
	0.1549	0.1563	0.1487	0.1282	0.1219	0.0840	0.0714
$\mu_1 - \mu_2$	0.1352	0.1521	0.1383	0.1355	0.1219	0.0888	0.0713
	0.1690	0.1541	0.1560	0.1371	0.1285	0.0902	0.0812
	0.1589	0.1574	0.1574	0.1373	0.1373	0.0912	0.0912
	0.1477	0.1438	0.1468	0.1478	0.1342	0.1186	0.0837
	0.1441	0.1456	0.1568	0.1355	0.1292	0.1123	0.0832
	0.1549	0.1381	0.1588	0.1196	0.1222	0.0989	0.0684
$\mu_1 - \mu_3$	0.1458	0.1169	0.1336	0.1185	0.1219	0.1092	0.0784
	0.1645	0.1424	0.1523	0.1358	0.1295	0.1087	0.0771
	0.1589	0.1589	0.1574	0.1574	0.1373	0.1373	0.0912
	−0.0012	0.0711	−0.0104	0.0665	0.0007	0.0586	−0.016
	−0.0066	0.0571	0.0049	0.0591	0.0017	0.0448	0.0020
	0.0065	0.0143	0.0082	0.0167	−0.0005	0.0276	−0.0081
$\mu_2 - \mu_3$	0.0071	0.0121	−0.0061	0.0260	0.0011	0.0396	0.0077
	−0.0043	0.0129	0.0017	0.0289	0.0006	0.0317	−0.0046
	0.0000	0.1574	0.0000	0.1373	0.0000	0.0912	0.0000

timators involving the more inferior treatments tends to overestimate the true values.

5. Discussion. A simple sequential procedure for the comparison of several treatments has been evaluated in this paper with various data-dependent allocation rules. In particular, we have seen that the allocation rule has negligible effect upon the probability of choosing the best treatment at the end of the trial. Further, it was shown that in some cases the numbers of patients on the inferior treatments can be reduced below those for equal randomization, and at the same time, the number of patients in the trial can be reduced. The results in Section 4.3 show that the bias and variance for estimated treatment differences are little affected by the allocation rule used. Of course, as shown in Tables 1 and 2, the use of a highly adaptive rule, although significantly reducing the number of patients on the inferior treatments, can lead to a substantial increase in the ASN.

Our simulation results indicate that a simple allocation rule such as

TABLE 3B

Results for variance of $x_i(N) - x_j(N)$, *where* $N = \min(N_i,\ N_j)$, *when* $k = 3$: *upper figures in order are simulated values for Equal, JJT, Gittins, Hayre and Unequal; lowest figure is approximate value.*

μ_1	1.0	1.0	1.0	1.0	1.0	1.0	1.0
μ_2	0.0	0.5	0.5	0.75	0.75	0.875	0.875
μ_3	0.0	0.0	0.5	0.5	0.75	0.75	0.875
	0.2045	0.1276	0.1358	0.1209	0.1177	0.1331	0.1386
	0.1962	0.1285	0.1280	0.1152	0.1177	0.1303	0.1324
	0.2015	0.1316	0.1300	0.1175	0.1254	0.1255	0.1287
$\mu_1 - \mu_2$	0.1881	0.1285	0.1194	0.1146	0.1139	0.1252	0.1250
	0.2106	0.1274	0.1339	0.1159	0.1211	0.1370	0.1426
	0.2095	0.1330	0.1330	0.1181	0.1181	0.1344	0.1344
	0.1999	0.2067	0.1251	0.1434	0.1260	0.1258	0.1377
	0.1989	0.2105	0.1328	0.1344	0.1213	0.1278	0.1396
	0.2019	0.2083	0.1307	0.1434	0.1207	0.1209	0.1324
$\mu_1 - \mu_3$	0.1932	0.1980	0.1200	0.1318	0.1114	0.1274	0.1250
	0.2136	0.2129	0.1275	0.1424	0.1312	0.1279	0.1379
	0.2095	0.2095	0.1330	0.1330	0.1181	0.1181	0.1344
	0.2955	0.2475	0.2037	0.1866	0.1718	0.1600	0.1570
	0.3057	0.2557	0.2069	0.1768	0.1635	0.1586	0.1569
	0.3706	0.2852	0.2185	0.1938	0.1618	0.1447	0.1402
$\mu_2 - \mu_3$	0.3683	0.2893	0.2132	0.1880	0.1681	0.1550	0.1471
	0.3585	0.3039	0.2179	0.1928	0.1686	0.1565	0.1588
	0.1499	0.1330	0.1499	0.1181	0.1499	0.1344	0.1499

the generalized Hayre rule or the unequal randomization rule is an effective compromise between equal randomization and a highly adaptive rule such as the Gittins rule. For modest treatment differences, both of these randomized rules reduce the numbers of patients on the inferior treatments at the cost of only a small increase in the overall trial size.

We have considered one form of bias in this paper, namely estimation bias. Of course, when one allows the use of data-dependent allocation, there is also the possibility of selection bias. This can occur when the experimenter knows which treatment the next patient is more likely to receive, and he or she allows this to influence the type of patient admitted. This can lead to misleading conclusions at the end of the trial. Some simple illustrations of selection bias are described by Bather (manuscript in this volume).

The estimation results in Section 4.3 indicate that our approximations for the bias and variance are reasonably accurate for estimated differences involving Treatment 1. This was also shown by Bather and Coad (1992) for Bernoulli responses when using equal randomization. Of course, these approximations were derived by assuming that only

two treatments are being used. It would be interesting if we cold obtain improved approximations by considering all the treatments in the trial. We would then hope to provide reasonable approximations for the bias and variance of estimators involving two inferior treatments. Some related work by Siegmund (1993) and Betensky (1992) dealing with the comparison of three treatments may provide a basis for work along these lines.

Acknowledgments. The author is grateful to the Mathematical Sciences Division of National Science Foundation for providing support to attend the Joint Summer Conference on Adaptive Designs, and to the referees for several valuable comments which helped to improve the paper.

References

BATHER, J.A. and COAD, D.S. (1992). Sequential procedures for comparing several medical treatments. *Sequential Analysis* **11** 339-376.

BETENSKY, R.A. (1992). *A Study of Sequential Procedures for Comparing Three Treatments*, Stanford University (doctoral dissertation).

COAD, D.S. (1991). Sequential tests for an unstable response variable. *Biometrika* **78** 113-21.

COAD, D.S. (1994). Estimation following sequential tests involving data-dependent treatment allocation. *Statistica Sinica* **4** 693-700.

GITTINS, J.C. (1989). *Multi-Armed Bandit Allocation Indices.* Chichester: John Wiley.

HAYRE, L.S. (1979). Two-population sequential tests with three hypotheses. *Biometrika* **66** 465-74.

JENNISON, C., JOHNSTONE, I.M. and TURNBULL, B.W. (1980). Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means, School of Operations Research and Industrial Engineering, Cornell University (technical report).

JENNISON, C., JOHNSTONE, I.M. and TURNBULL, B.W. (1982). Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. *Proceedings of the Third Purdue Symposium on Statistical Decision Theory and Related Topics*, (S.S. Gupta and J. Berger, eds). New York: Academic Press, 55-86.

ROBBINS, H. and SIEGMUND, D.O. (1974). Sequential tests involving two populations. *Journal of the American Statistical Association* **69** 132-39.

SIEGMUND, D. (1993). A sequential clinical trial for comparing three treatments. *The Annals of Statistics* **21** 464-83.

TURNBULL, B.W., KASPI, H. and SMITH, R.L. (1978). Adaptive sequential procedures for selecting the best of several normal populations. *Journal of Statistical Computation and Simulation* **7** 133-50.

WOODROOFE, M. (1989). Very weak expansions for sequentially designed experiments: linear models. *The Annals of Statistics* **17** 1087-102.

MATHEMATICS DIVISION
UNIVERSITY OF SUSSEX
FALMER, BRIGHTON, BN1 9QH
UNITED KINGDOM